Supplementary Table S15. Predictions of the effect of substitutions in human *NAT1* and *NAT2* coding sequences according to PolyPhen, SIFT and PANTHER cSNP Scoring compared to the phenotype assignments of the official nomenclature of human *NAT* alleles (http://nat.mbg.duth.gr).

A description of the scores and predictions returned by each of the three online software tools is provided in Supplementary File S1.

	Substitution		Phenotype assignment	PolyPhen		SIFT		PANTHER cSNP Scoring	
Haplotypes	cDNA	protein	(http://nat.mbg.duth.gr)	Score[1]	Prediction[2]	Score[3]	Prediction[4]	PSEP[5]	Prediction[6]
NAT1									
*NAT1*4* (reference)									
*NAT1*17*	C190T	R64W	Lower than *NAT1*4*	1 (0.00 -1.00)	PRD	0 (3.07, 81)	A	4200	PRD
*NAT1*22*	A752T	D251V	Lower than *NAT1*4*	1 (0.00-1.00)	PRD	0 (3.07, 80)	A	455	PRD
*NAT1*14B*	G560A	R187Q	Lower than *NAT1*4*	0.772 (0.85-0.92)	POD	0.54 (3.07, 81)	T	455	PRD
*NAT1*21*	A613G	M205V	Equivalent to *NAT1*4*	0 (1.00-0.00)	B	1 (3.07, 81)	T	91	B
*NAT1*24*	G781A	E261K	Equivalent to *NAT1*4*	0.008 (0.96-0.76)	B	0.03 (3.07, 80)	A	455	PRD
*NAT1*25*	A787G	I263V	Equivalent to *NAT1*4*	0 (1.00-0.00)	B	0.95 (3.07, 80)	T	220	POD
*NAT1*30*	G445A	V149I	Unknown	0 (1.00-0.00)	B	1 (3.07, 81)	T	6	B
NAT2									
*NAT2*4* (reference)			Rapid						
*NAT2*12A*	A803G	K268R	Rapid	0.011 (0.96-0.78)	B	0.23 (3.24, 51)	T	176	B
*NAT2*19*	C190T	R64W	Slow	1 (0.00-1.00)	PRD	0 (3.08, 51)	A	4200	PRD
*NAT2*14A*	G191A	R64Q	Slow	1 (0.00-1.00)	PRD	0 (3.18, 53)	A	4200	PRD
*NAT2*5D*	T341C	I114T	Slow	0.359 (0.90-0.89)	B	0.04 (3.18, 53)	A	220	POD
*NAT2*17*	A434C	Q145P	Slow	1 (0.00-1.00)	PRD	0 (3.18, 54)	A	456	PRD
*NAT2*6B*	G590A	R197Q	Slow	1 (0.00- 1.00)	PRD	0.01 (3.18, 54)	A	842	PRD
*NAT2*18*	A845C	K282T	Rapid	0.999 (0.14-0.99)	PRD	0.03 (3.39, 46)	A	220	POD
*NAT2*7A*	G857A	G286E	Slow[7]	0.317 (0.90-0.89)	B	0.62 (3.69, 42)	T	31	B
*NAT2*10*	G499A	E167K	Slow[7]	0.001 (0.99-0.15)	B	1 (3.18, 54)	T	176	B
*NAT2*13B*	C578T (C282T)[8]	T193M (Y94Y)[8]	Unknown[9]	1 (0.00-1.00)	PRD	0 (3.18, 54)	A	456	PRD

[1] PolyPhen score : probability that a substitution is damaging; sensibility and specificity in brackets.

[2] PolyPhen prediction : "benign" (B), "possibly damaging" (POD), "probably damaging" (PRD).

[3] SIFT score : probability that a substitution is tolerated; median sequence information and number of sequences used for the prediction in brackets.

[4] SIFT prediction : T: “tolerated” (T), A: “affect protein function” (A).

[5] PANTHER cSNP Scoring PSEP (position-specific evolutionary preservation) : length of time (in millions of years) of preservation of a position.

[6] PANTHER cSNP Scoring prediction : "probably damaging" (PRD), "possibly damaging" (POD), "probably benign" (B).

[7] The official nomenclature of human *NAT2* alleles reports the phenotype as “Slow, Substrate dependent ?”.

[8] The official nomenclature of human *NAT2* alleles has synonymous SNP C282T (Y94Y) as signature mutation of all *NAT2*13* haplotypes.

[9] No phenotype assignment provided in the official nomenclature of human *NAT2* alleles; the substitution has been predicted (with in-silico tools) as damaging by (Patin et al. 2006b).

Reference

Patin E, Harmant C, Kidd KK, Kidd J, Froment A, Mehdi SQ, Sica L, Heyer E, and Quintana-Murci L. 2006. Sub-Saharan African coding sequence variation and haplotype diversity at the NAT2 gene. Hum Mutat 27(7):720.